Year: 2008

Assisting Human Cognition in Visual Data Mining

Simoff, Simeon J ; Böhlen, Michael Hanspeter ; Mazeika, Arturas

Abstract: As discussed in Part 1 of the book in chapter Form-Semantics-Function. A Framework for Designing Visualisation Models for Visual Data Mining the development of consistent visualisation techniques requires systematic approach related to the tasks of the visual data mining process. Chapter Visual discovery of network patterns of interaction between attributes presents a methodology based on viewing visual data mining as a reflection-in-action process. This chapter follows the same perspective and focuses on the subjective bias that may appear in visual data mining. The work is motivated by the fact that visual, though very attractive, means also subjective, and non-experts are often left to utilise visualisation methods (as an understandable alternative to the highly complex statistical approaches) without the ability to understand their applicability and limitations. The chapter presents two strategies addressing the subjective bias: guided cognition and validated cognition, which result in two types of visual data mining techniques: interaction with visual data representations, mediated by statistical techniques, and validation of the hypotheses coming as an output of the visual analysis through another analytics method, respectively.

DOI: https://doi.org/10.1007/978-3-540-71080-6_17

Originally published at:
Simoff, Simeon J; Böhlen, Michael Hanspeter; Mazeika, Arturas (2008). Assisting Human Cognition in Visual Data Mining. In: Simoff, Simeon J; Böhlen, Michael Hanspeter; Mazeika, Arturas. Visual Data Mining: Theory, Techniques and Tools for Visual Analytics. Berlin / Heidelberg: Springer, 264-280.
DOI: https://doi.org/10.1007/978-3-540-71080-6_17

Assisting Human Cognition in Visual Data Mining

Simeon J. Simoff[1,2], Michael H. Böhlen[3], Arturas Mazeika[3]

[1]School of Computing and Mathematics
College of Heath and Science
University of Western Sydney, Australia
s.simoff@uws.edu.au
[2]Faculty of Information Technology
University of Technology, Sydney, Australia
[3]Faculty of Computer Science
Free University of Bozen-Bolzano, Italy
{boehlen, arturas}@inf.unibz.it

Abstract. As discussed in Part 1 of the book in chapter "Form-Semantics-Function – A Framework for Designing Visualisation Models for Visual Data Mining" the development of consistent visualisation techniques requires systematic approach related to the tasks of the visual data mining process. Chapter "Visual discovery of network patterns of interaction between attributes" presents a methodology based on viewing visual data mining as a "reflection-in-action" process. This chapter follows the same perspective and focuses on the subjective bias that may appear in visual data mining. The work is motivated by the fact that visual, though very attractive, means also subjective, and non-experts are often left to utilise visualisation methods (as an understandable alternative to the highly complex statistical approaches) without the ability to understand their applicability and limitations. The chapter presents two strategies addressing the subjective bias: "guided cognition" and "validated cognition", which result in two types of visual data mining techniques: interaction with visual data representations, mediated by statistical techniques, and validation of the hypotheses coming as an output of the visual analysis through another analytics method, respectively.

Introduction

As vision is by far the human's most powerful input information channel, many researchers in the area of data mining and computing science have worked on information visualisation methods that facilitate human understanding of data and data analysis results. Humans construct such understanding by forming a mental model which captures only the essence of the phenomena in consideration. In terms of visual data mining this means that humans do not necessarily need detailed visualisation of the whole data set, but a considerably lower amount of information, generated out of that data set, which, when visualised, is sufficient in forming human perception and model of the patterns in the original data set. In other words, to be able to capture the essence of the data, visualisation techniques should "fit" human cognition, i.e. the human mental model. The consistent development of such visualisation techniques and the corresponding visual data mining models remain key research issues in the area. The first part of this chapter addresses these issues. We look at visual data mining techniques that aim at guiding human cognition by abstracting and manipulating visual forms from the row data. The algorithms embed statistical procedures for composing the visual presentations in a way that ensures the "objectiveness" of observed patterns. The second part of the chapter looks at the ways of validating patterns that are discovered visually from raw data. As the ultimate goal of the knowledge discovery process is in gaining deeper understanding of the phenomenon, "seeing" data assists in making sense out of it. However, "seeing" is a subjective process and depends on the mapping of data attributes to the dimensions of the visual presentation. An issue in visual data mining is how to validate the observed pattern. For example, an obvious clustering pattern visible in 2D or 3D projections of the data points may not necessarily hold in other projections, or when considering all projections, and yet, it may hold the key to understanding the structure of the data. The second part of the chapter is devoted to this issue.

Visual bias in visual data mining

As an approach that integrates the exploration and pattern spotting abilities of the human mind with the processing power of computers to form a powerful knowledge discovery environment, visual data mining is expected to capitalise on the best of both worlds [1-3]. Visual data mining methodologies and systems offer functionality that characterises structures and displays data, targeting human capabilities that perceive patterns, anomalies, relationships, and tendencies. The utilisation of visual communication between computer and the human for depicting novel and interpretable patterns in data has been emphasised in the various definitions of visual data mining. In this chapter we recall the three ways of embedding visualisation in the data mining process [2], which have been extended in chapter "Complementing visual data mining with the sound dimension: Sonification of time dependent data". In this chapter we look at the visual data mining from a cognitive perspective, considering the subjective aspects of the processes and how to enable the objectivity of the outcomes. These issues are important, regardless of whether visualisation is applied only to the final output (results) of the data mining algorithms or to the raw data and intermediate output of data algorithms. As the aim is to condense the information and to enable its interpretation, different visual representations of the same information target different cognitive styles. Cognitive style is a term introduced by cognitive psychology to describe the preferred way individuals process information, i.e. a person's typical mode of thinking, perceiving, remembering and utilising information in what s/he does [4]. Individual differences in abilities ("individual style") are described in terms of peak performance over unipolar dimensions, with ranking scales ranging from zero to a maximum value. Consequently, one can compare two individuals in having "more" of the one ability and "less" of another. Cognitive styles are usually considered to be bipolar dimensions. The combination of such dimensions usually denotes a tendency to process information and behave in a certain manner.

As the pattern discovery in visual data mining is a form of a decision making process in order to make sense of what is seen, understanding and taking in account the cognitive aspects of such process are central to enabling successful outcomes. In the earlier works Eduard Tufte [5, 6] provides an impressive argument for the enabling power of proper data and information visualisation when it comes to support our cognition in decision-making. Card, Mackinlay and Shneiderman [7] visual displays of data and abstract information stimulate and amplify cognition through a knowledge crystallisation process [an analogy can be drawn from the design fixation process coming from studies of how designers work]. Spence [8] argues that as a result of the interaction with the visual representation of the data and information, the analyst develops a mental model (cognitive map); a cyclical process of human activity creates and recreates this mental model. Tufte [9] discusses the links between cognitive styles and the styles of information display. Subjects with high spatial abilities have an inclination towards constructing spatial or abstract virtual worlds when interpreting abstract data [10]. [10] also distinguished between the ventral stream (the "what" pathway), which plays major role in object identification (shape and colour), and the dorsal stream (the "where" pathway), which plays major role identifying spatial relations among objects. The dorsal and ventral streams are claiming responsibility for constructing internal representations in the absence of external representations. These differences are important when considering ways for assisting the visual data mining process. For example, depending on human cognitive style, associations may be easier for comprehension through a visualisation of association rules [11], rather than through a list of rules sorted by confidence, support, lift and other measures of coverage and "interestingness").

When visual data mining operates with raw data input is included in the loop, as shown in the analyst can have a full picture of the data space and perform investigation without any assumptions or preconceptions. Depending on the case the analyst may interpret the visual patterns and the analysis may be completed without running additional data mining algorithms. When for explorative analysis this step may be sufficient, predictive and classification tasks require backup with analytical methods.

On the other hand, though intuitive and attractive, displaying the raw data may not necessarily lead to successful results for number of reasons. Due to the limits of visual displays, we usually are interacting with 2D or 3D projection(s) of the multidimensional data set. Such projections may not necessarily present a cross-section that is the best match for our cognition, as shown in the example of visualising a large Web log data set in Fig. 1. As a result, the visual projection may introduce a visual bias – a perceptual bias of the visual attention leading to selective rendering, which may mislead the analyst. This may result in either:

1. missing in spotting structures which happened to be encoded in the data set, but not necessarily well projected in the selected visual representation of the data set, or/and
2. invalid interpretation of the visual display.

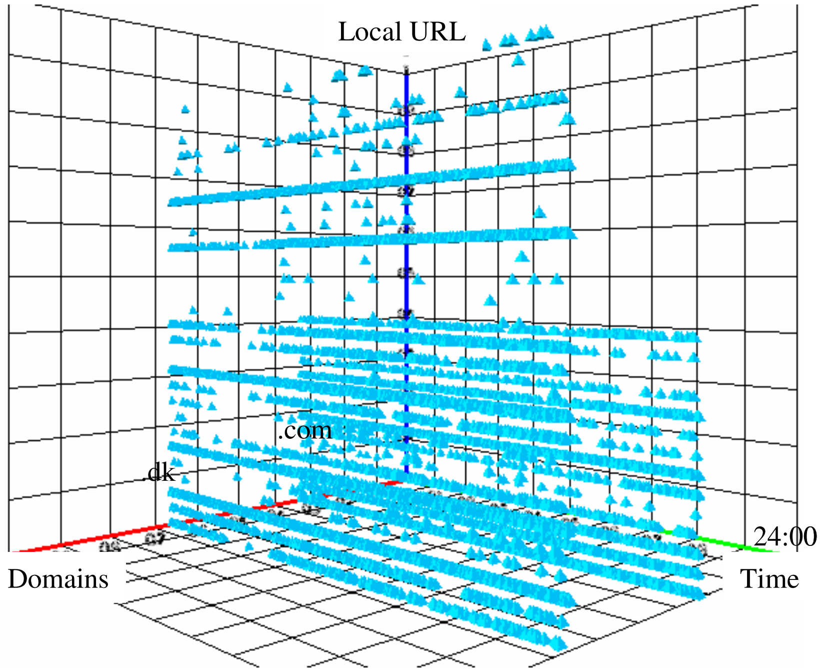


Fig. 1. Raw data does not necessarily reveal its structure readily.

These two different outcomes are addressed with two different methods and their supporting algorithms. We argue that the display of raw data should be equipped with algorithms that assist our perception system to depict forms and structure. Such procedures include *algorithms for abstracting forms from raw data* and *algorithms for manipulating feature values* to assist in depicting visual patterns. The first group of algorithms constitute the method of *guided cognition* in data exploration as they use statistical characteristics to ensure that the form abstraction is accurately visualising the structure of the data. In terms of visual data mining techniques this results in interactions with visual data representations, that are mediated by statistical techniques, i.e.

guided cognition = visualisation + statistical technique

The second group of algorithms constitute the method of *validated cognition*, as it requires additional validation techniques. In terms of visual data mining techniques this results in generating hypotheses through identifying visual patterns in the data during visual exploration and then validating these hypotheses through another analytics method, i.e.

validated cognition = visual exploration + analytics technique

Both methods rely on the tight integration of the visual data mining techniques with other procedures that assist in validating the results. They will require a number of intermediate visualisation steps. This includes, but is not limited to the "visualisation of intermediate result" as described in [2]. The interaction between visualisation and data mining algorithms is in broader sense than visualisation of algorithmic decisions in the tightly integrated visualisation, discussed in [12].

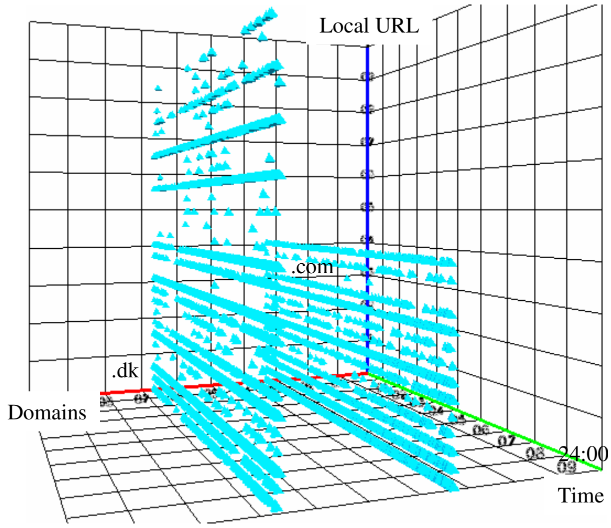
Addressing the visual bias in visual data mining

We illustrate the ideas underpinning the two methods on portions of two case studies, one with a large and one with a small data set.

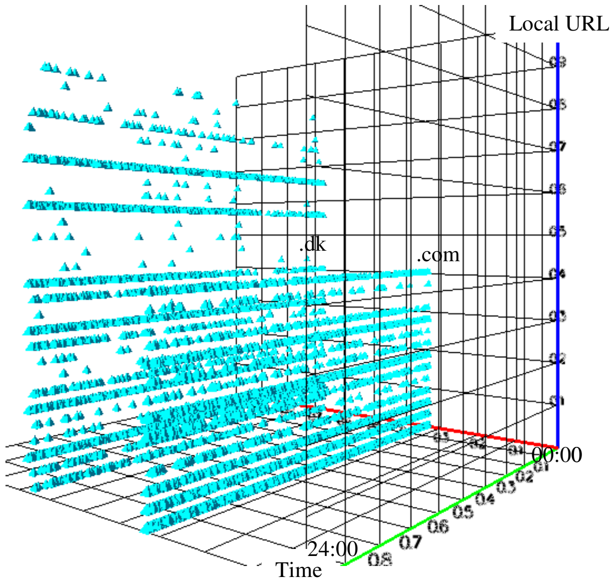
The method of guided cognition, implemented through embedded statistical techniques

This section addresses the issue of stimulating ("guiding") our cognitive mechanisms to discover visual forms that lead to meaningful interpretation. We use the density surface technique [13] and the change of density level to guide the data miner through the process. The data source is a three dimensional web log. The first dimension is the domain (we aggregated the URL attribute to the last part of the server name). There is a variety of the activity from different domains, many of them with very few clicks only. In our plots we show the two most frequent domains: .com and .dk (the local domain). The second dimension is the date and time. For our analysis we picked a 24 hour segment starting at 00:00. The third dimension is the local Web page that was requested. The organisation is such that the general pages (departmental home page, high level departmental descriptions of teaching, research, etc) are towards the beginning of the axis whereas the more specialised pages (personal pages, specific research projects, courses, etc) are towards the end of the axis.

As the methods that we propose are applicable to different underlying visualisations, we have selected one of the most common types of visual representations – the scatter plot, in order to illustrate how the method works. Fig. 2 shows a scatter plot visualization of the raw data. In our case each triple of the data set, e.g. (.dk, 08:53:12, www.cs.auc.dk/research/), is represented as a primitive visual object (a tetrahedron in our specific examples). Clearly the visualizations in Fig. 2 provide limited information only. We see that some pages are requested very frequently. This is a very dominant pattern and yields a visualization with pronounced lines. While true it is also information that we most likely knew already before. It is also very difficult to recognize differences between the lines. While it is possible to tell apart pages that are hardly ever accessed from pages that are accessed often it is not possible to say if ".com" and ".dk" access the same pages most often. In general the interpretation of a 3D scatter plot benefits a lot from rotations. In terms of interaction with scatter plot visuals, rotation is one of the preferred operations. Rotation alleviates the problems of occlusion and corroborates the perception of the 3D space.



a.

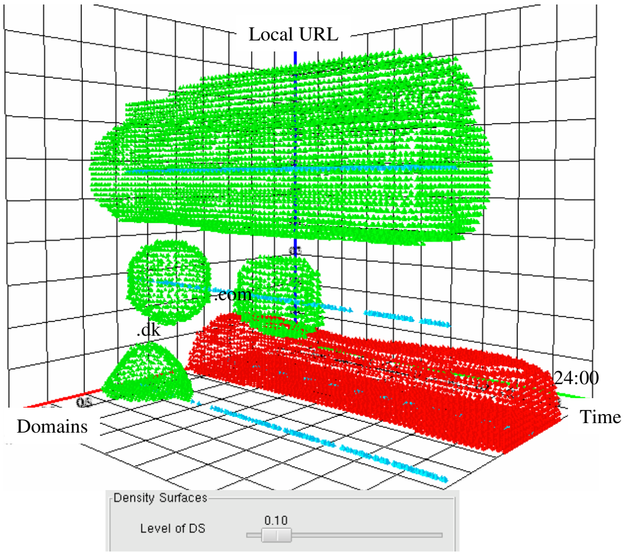


b.

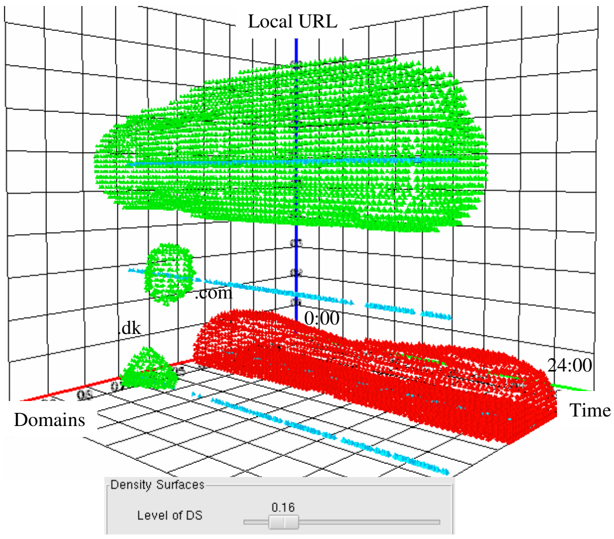
Fig. 2. Simple operations with the visual perspective, such as rotation rarely can help in uncovering the patterns encoded in the raw data.

The change in the views in Fig. 1 and Fig. 2 illustrates the improvements of the visual perspectives. However, with large data sets it may not necessarily assist in the depiction of the structures embedded in the data. The two improved in terms of occlusion projections in Fig. 2a and Fig. 2b emphasise this problem – after all rotation steps an analyst can hardly say much more from what s/he could say after exploring the projection of data set in Fig. 1.

The method of guided cognition utilises statistical transformations and models in order to enrich the semantics of the visualisation in terms of patterns in the data. Fig. 3 shows the result of complementing the data set, visualised in Fig. 2, with model information.



a. Patterns of sequential access emerged at the level of density around 0.10



b. Further increase of the density level leads to clearer isolation of emerging patterns

Fig. 3. “Switching on” the density surface technique reveals some patterns

In this case the model information is a density surface. Roughly, the density of a 3D data set is a 4D structure and we visualize a cut at a specific density level. In the context of supporting the visual discovery process, the level value is controlled smoothly in order to depict the range of levels when patterns start to emerge, then become clearly distinguishable and stable. These cuts are the surfaces displayed in Fig. 3. A density surface is visualized using a very simplistic method: we draw points on the surface as primitive objects. This is a simplistic but effective method as illustrated in the sequence of steps in Fig. 3. Detailed visual investigation of the data set through progressive change in levels and the corresponding changes in the visual displays of the data set are shown in Fig. 3a, Fig. 3b and Fig. 4.

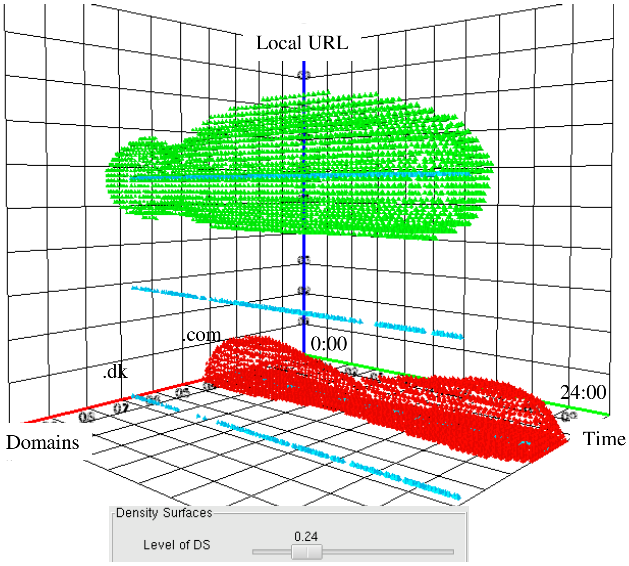


Fig. 4. Further increase of the density level leads to isolation of the final destinations

The process shown in Fig. 3 and Fig. 4 reveals information that we could not see from looking at the scatter plots. It becomes clear that .dk and .com users access quite different pages. Users from .com (the dark surfaces) most frequently access the departmental home page and possibly a few other pages from there. Users from .dk (the light surfaces) access the home page much less. Instead they directly proceed to the pages about teaching and research. By varying the density level from 0.16 to 0.08 we progress our understanding of the distribution of the clicks.

Fig. 4 and Fig. 5 show how the data set can be explored further by varying the density levels. We get an in-depth overall understanding of the distribution of the accessed pages. As there is no single perspective that works best in all cases, throughout the investigation rotation is used to view the surfaces from different perspectives. As a result of the clear separation of the patterns, rotation is again a useful interactive operation in this visual data mining process.

The above discussed examples clearly demonstrate the need for embedding statistical techniques that enhance the raw data with model information to guide visual analysis. The next case study illustrates an opposite approach where the visual analysis is done without embedded statistical technique, over the raw data. The hypothesis formulated this way are then validated (hence the name "validated cognition") using an analytical technique.

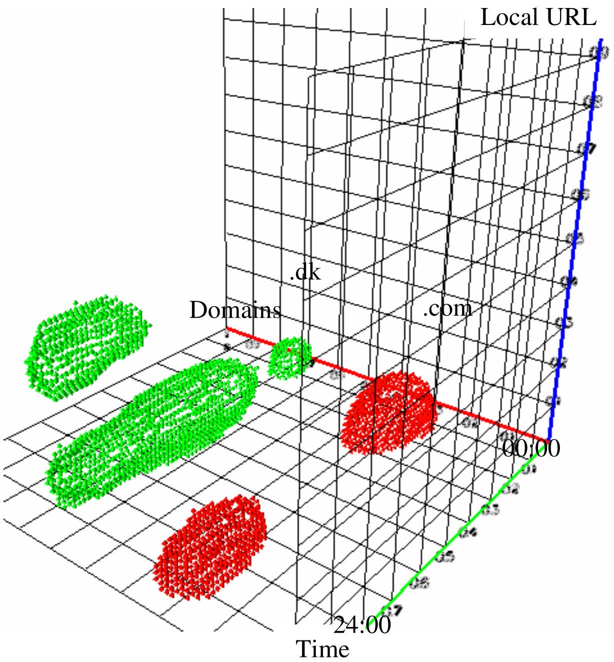


Fig. 5. "Rotation" can now assist in revealing the different structures in the data set.

The method of validated cognition, implemented through a combination of visual data mining techniques

The problem, which the method of "validated cognition" addresses, perhaps dates back to the works of John W. Tukey. Tukey was looking at methods for robust analysis in the presence of violation of initial assumptions, including robust visualization methods. He emphasised that seeing may be believing or disbelieving, but above all, data analysis involves visual, as well as statistical, understanding. Perhaps the most famous and certainly one of the oldest visual explanations in mathematics is the visual proof of the Pythagorean Theorem. This proof is unusual in its brevity and its complete appropriateness to the problem. Thus, the rational is that a visual reasoning approach to data mining may be able to overcome some of the difficulties experienced in the comprehension of the information encoded in data sets and the models derived by pure analytical methods. These issues have been extensively discussed in the consecutive workshops on visual data mining [14-17].

What remained untouched was the issue of validating the results of our perception of the visualised forms.

The method of validated cognition proposed in this chapter is described using a case study of a company, whose managers wanted to grasp a better and valid picture of their social capital. In particular, when running development projects, for example, software engineering projects, project managers may like to collect intelligence about the project team, including who are the informal leaders, the attributes that describe the team, the experts, the good executors and other characteristics of the team. The executives of any organisation can look for the proactive experts and idea generators among their employees, who emerged during the different projects. In our case study the data source includes the email interactions between company employees, record of documents flow, time-stamped consecutive versions of these documents, plus demographic data – a data set similar in structure to the Enron data set[1] . For the illustration of the methodology discussed in this chapter, we limit our investigation to the email data and will be using the email messages only in a half a year window, observing the communication of a group of employees that work on a globally distributed project.

The goal of the study was to identify team members that emerged out of the group during the project run, based on their knowledge, expertise and attitude, and to identify the attributes that characterise their communication behaviour. To ensure that the project runs in a truly distributed team environment, project members were recruited from company sections in different regions of the globe, so that no two project members were co-located. Team members were asked to communicate via e-mail, both for individual and group communications (group messages were sent to a common listserv account) in order to enable reliable data collection. Fax and phone were excluded as communication means. As part of the company some team members knew each other, but they had no history of working with each other as a team. For the illustration of the method, we use only the email communication data set collected over a 3 month period. It includes 2954 messages, out of which, 1489 email messages were exchanged between individuals, the rest were sent to the project list. The assumption is that proactive participants will have more intensive communication exchanges in terms of individual message frequencies (both generating and being addressed) and by the amount of information that they communicate.

In this example, we illustrate a two stage analysis. The first stage is a visual analysis of individual actors based on the attributes used to describe them. We use four attributes "Interaction" (all the outgoing messages for particular individual), "Addressed" (all messages addressed to particular individual), "Words" (the total amount of words in all outgoing messages) and "Words (mean)" (the average amount of words in a message) We explore the structure of the data set and identify potential candidates for proactive team members. As we expect their behaviour to distinguish them from the rest of the group, we look for outliers or/and small isolated clusters. On the second stage of the analysis, the results of the initial selection are validated by running a network analysis of the communication network established between project participants.

[1] http://www.cs.cmu.edu/~enron/

Visual analysis

The visualization of the structure of the data set revealed that a large group of participants had a relatively low intensity of individual communication and they cluster closer to the origin along the attributes “Interaction”, Addressed and “Words”. They were removed from the visualisation. The data for the remaining 30 project members is shown in Fig. 6[2].

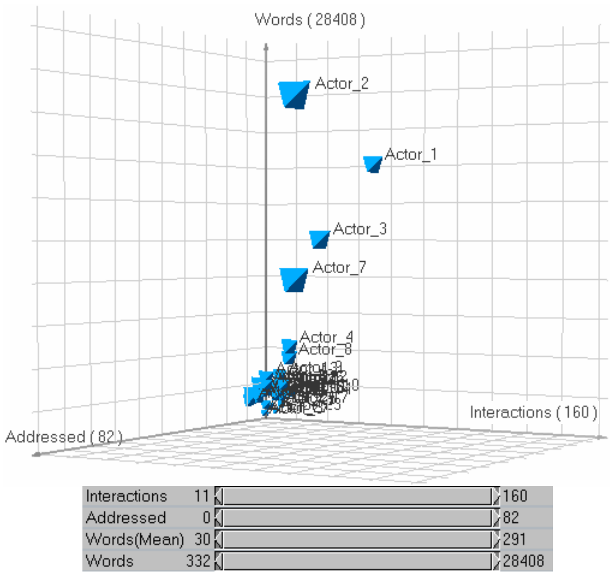


Fig. 6. Exploring the structure of the project group data set

Each pyramid represents a projection of a data point, where the coordinates of the centre in our case are defined by the values of the above mentioned attributes. Project participants are labeled as “actors”. Actors 1-3 and 7 are candidates for consideration, as they are out of the main clusters of project members. Then these points are filtered out. Actors 4 and 8 need further clarification. The attributes were remapped to the visual features and one dimension switched off. As a result, actor 5 emerged and was included in the candidate set, as shown in Fig. 7. Fig. 8 illustrates how actors 4 and 6 were depicted. In a similar way, using filtering, remapping of attributes and navigation operations, actors 9 and 13 were identified. The final candidate list includes 1-6, 9 and 13.

[2] For the visual analysis in this case we used the off-the-shelf product Miner3D - a visual data mining tool by Dimension 5 (http://www.miner3d.com).

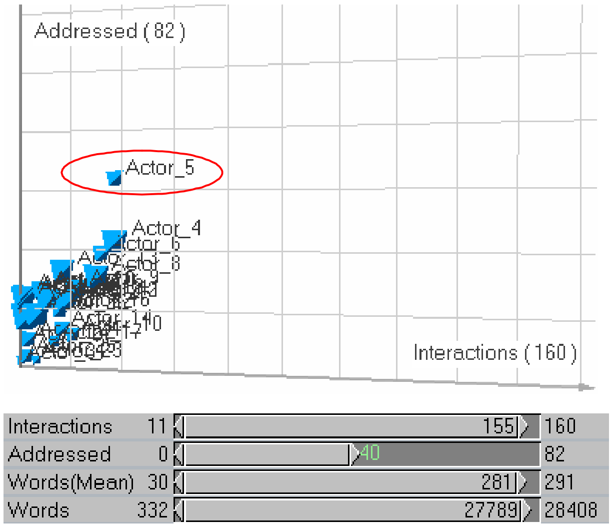


Fig. 7. Actor 5 emerged and was included in the candidate set.

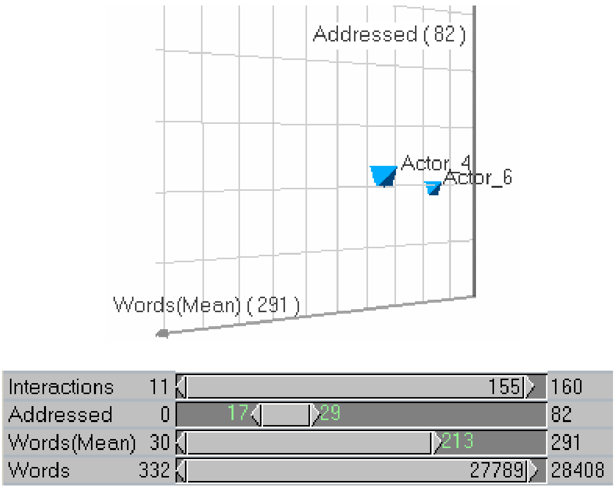


Fig. 8. Projection which depicted Actor 4 and Actor 6.

Validation

The aim of the network analysis is to refine and validate the candidate list based on the differences in the position of different people in the project team as the project unfolds. We analysed the interactions between individuals during the project. The data set for this analysis included only the records in an email's “To:” and “From:”

fields. The "Subject:" field and the actual content of the email messages at this stage were ignored. A large group of participants had a relatively low intensity of individual communication (the filtering threshold was set to 10 messages), which left 30 project members in consideration. The network structure derived from the interaction between these project members is shown in Fig. 9[3].

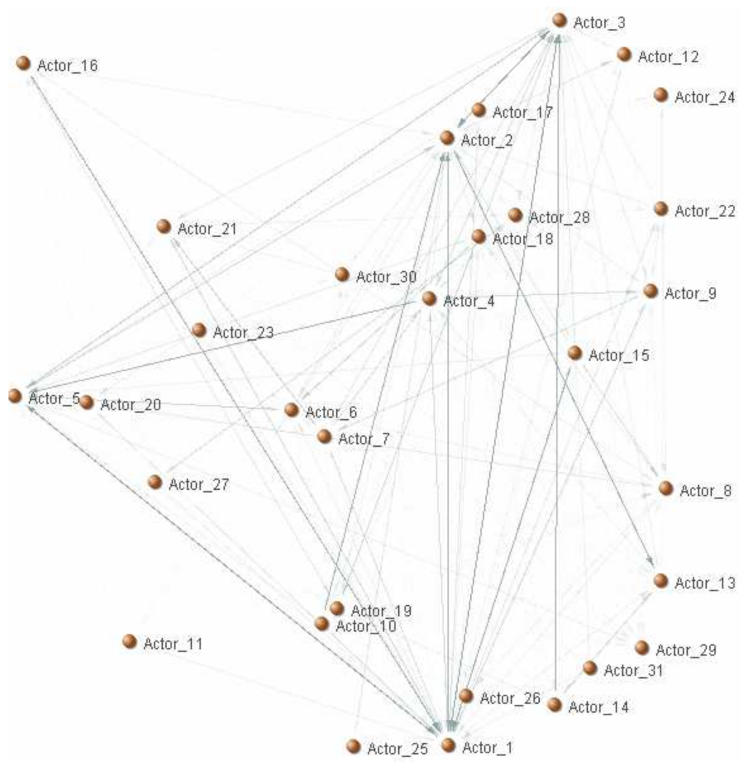


Fig. 9. Structure of the interactions between project members

The network is non-uniform and directed. Network parameters are presented in Table 1. The diameter of a network – the length of the longest geodesic path to be found in that network [geodesic path is the shortest path between two nodes], indicates that in our case, the network is of the "small world" type [18], i.e. participants are not far from each other, in terms of message passing. This is confirmed by the low value of average eccentricity of nodes in the network (eccentricity summarises how far a node is from the node most distant from it in the network). The density of our network [the ratio of the total number of edges to the number of all possible edges in that network] shows that it seems to be of the "scale free" type, where some hubs form the key nodes, with a number of "weakly" connected nodes. The cohesion index [the ratio of the number of mutual connections

---

[3] We use Agna 2.1.1 and UCINET 6 to conduct the network analysis.

in the network to the maximum possible number of such connections] shows a relatively low amount of bidirectional interactions.

Table 1. Network parameters

Nodes	Edges	Diameter	Eccentricity (average)	Density	Cohesion
31	197	4	3	0.21	0.11

Consequently, we look at the sociometric statistics of the individual nodes, to select potential emergent authorities. These statistics for the thirty project members are shown in Table 2. The indicators in the table include the emission degree of each node [the sum of all values corresponding to the edges originating in that node], the reception degree of each node [the sum of all values corresponding to the edges incident (directed) to that node] and the sociometric status of each node [the sum of its reception and emission degrees, relative to the number of all other nodes]. Setting up a reasonable threshold of 10 for the first two statistics and 0.5 for the sociometric status we identify the nodes of interest that include actors 1-5, 8, 9, 13 and 21. This confirmed the initial list of actors (except, actor 6) generated during the visual stage. Actor 6 is a border case (as seem to be actor 7), and will need further separate investigation. We got in actor 8, which was originally spotted as a possible candidate, but was not identified through further visual manipulations. The final list, refined by content analysis of communication messages, included all actors, whose rows are shaded with grey in Table 2. The details of this analysis are beyond the scope of the chapter.

The above presented approach showed that for smaller data sets, choosing an appropriate mapping of the attributes to the features of the visual mining system can allow to formulate some hypothesis about a phenomenon encoded in the data. However, these hypothesis need to be validated and refined separately with an analytical technique. The approach is well-suited for mining of integrated data sets, where for example, communication data is only part of the data set. To some extent, the method of validated cognition can be viewed as an analogy of the triangulation method in qualitative research. It also requires selection of the methods involved in the validation step for the data set in consideration, of the actual process. The combination of the techniques requires consideration of the visual characteristics of the techniques. CEDA methodology [19], subsequently enhanced in CEMDA [20] can be helpful in combining the validation methods. Originally derived for the area of Internet research, where the replication of experiments is almost impossible due to the nature of the network protocols, CEDA approach has provided a framework for consistent integration of different qualitative and quantitative techniques.

Table 2. Distribution of sociometric statistics of individual project members

Node	Emission	Reception	Status
Actor_1	46.0	61.0	3.67
Actor_2	28.0	52.0	2.67
Actor_3	33.0	40.0	2.43
Actor_4	18.0	16.0	1.13
Actor_5	23.0	30.0	1.77
Actor_6	7.0	16.0	0.77
Actor_7	19.0	7.0	0.87
Actor_8	14.0	11.0	0.83
Actor_9	14.0	11.0	0.83
Actor_10	15.0	3.0	0.60
Actor_11	5.0	1.0	0.20
Actor_12	4.0	6.0	0.33
Actor_13	11.0	14.0	0.83
Actor_14	16.0	5.0	0.70
Actor_15	14.0	7.0	0.70
Actor_16	9.0	7.0	0.53
Actor_17	8.0	3.0	0.37
Actor_18	8.0	9.0	0.57
Actor_19	13.0	8.0	0.70
Actor_20	3.0	6.0	0.30
Actor_21	11.0	11.0	0.74
Actor_22	6.0	6.0	0.40
Actor_23	7.0	0.0	0.23
Actor_24	2.0	2.0	0.13
Actor_25	4.0	0.0	0.13
Actor_26	1.0	5.0	0.20
Actor_27	5.0	2.0	0.23
Actor_28	5.0	8.0	0.43
Actor_29	2.0	7.0	0.30
Actor_30	1.0	5.0	0.20

Summary and future directions

Visual data mining looks at having an access to the entire data set in its most granular level. As mentioned in chapter "Visual discovery of network patterns of interaction between attributes" earlier in this book, visual data mining can be viewed as a "reflection-in-action" process. We have illustrated two different methods fitting this view, namely "guided cognition" and "validated cognition". This work is motivated by the fact that visual, though very attractive, means also subjective, and non-experts are often left to utilise visualisation methods (as an understandable alternative to the highly complex statistical approaches) without the ability to understand their applicability and limitations, and the impact of these limitations on the interpretation of the results. It is questionable to what extent exploring raw data (especially large amounts in 3D systems) actually helps the mining as perceptual faculty is overloaded and confused, rather than stimulated. An important issue is the identification of particular ranges, beyond which only a visual mining technique with embedded guidance can help. Another issue encountered in this work is the lack of effective interactive summarisation models that can reduce the computational load in real-time reflective techniques.

Taking the overall discussion a one step further, in many cases in visual data mining finding patterns during an interactive session requires from the analyst a substantial change in perspective. In this context, one of the ultimate goal of cognitive support in visual data mining is to facilitate the analyst to get "outside of the box", i.e. out of the set expectations and assumptions based on background knowledge about the data set, that may introduce bias in the visual pattern evaluation. In terms of analytics, this means enable change of the conceptualisation the problem domain where the data come from. Both methods presented in this chapter support this process. The guided cognition approach enables the shifting of the thinking by emphasising the generation of visual forms that encode structures in the data, which are guaranteed by the mathematical properties of the technique used to generate these forms. In some sense, the validity of the observations is achieved during the discovery. The validated cognition approach takes a softer approach towards the validity of the visual patterns during their discovery, but puts them to the test in the second stage – through an analytics method which runs in hypothesis testing mode rather than a discovery mode. Hence, further work in this direction can look at:

- the development of guiding techniques (density surfaces is only on of these techniques)
- the development of algorithms for automatic selection of guiding techniques, i.e. the criteria for selection of different techniques;
- the development of algorithms for automatic selection of validation techniques.

Last, but not least, this research has been supported by the Australian Research Council. We would like to acknowledge the use of the 3DVDM engine for the first case and Miner3D (http://miner3d.com/) in the second case.

References

1. Wong, P.C., Visual Data Mining. *IEEE Computer Graphics and Applications*, 1999 (September/October): p. 1-3.
2. Ankerst, M., *Visual Data Mining*. Fakultät für Mathematik und Informatik. 2000, München: Ludwig-Maximilians-Universität.
3. Schulz, H.-J., T. Nocke, and H. Schumann. A Framework for Visual Data Mining of Structures, in *Proceedings of the Twenty-Ninth Australasian Computer Science Conference (ACSC2006)*. 2006. CPRIT Vol. 48, Hobart, Tasmania, Australia: Conferences in Research and Practice in Information Technology.
4. Witkin, H.A. and D.R. Goodenough, *Cognitive Styles: Essence and Origins*. 1981, New York: International Universities Press. 141.
5. Tufte, E.R., *Visual and Statistical Thinking: Displays of Evidence for Decision Making*. 1997: Graphics Press.
6. Tufte, E.R., *Visual Explanations: Images and Quantities, Evidence and Narrative*. 1997: Graphics Press. 156.
7. Card, S.K., J.D. Mackinlay, and B. Shneiderman, eds. *Readings in Information Visualization: Using Vision to Think*. 1999, Morgan Kaufmann.
8. Spence, R., *Information Visualization*. 2001: Addison-Wesley.
9. Tufte, E.R., *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*. 2003: Graphics Press.
10. Kozhevnikov, M., S. Kossyln, and J. Shephard, Spatial versus object visualizers: a new characterization of visual cognitive style. *Memory and Cognition*, 2005, **33** (4): p. 710-726.
11. Hofmann, H., A. Siebes, and A. Wilhelm. Visualizing association rules with interactive mosaic plots, in *Proceedings of ACM SIGKDD Int. Conf. On Knowledge Discovery and Data Mining (KDD 2000)*. 2000. Boston, MA: ACM Press.
12. Ankerst, M., G. Grinstein, and D. Keim. Visual Data Mining: Background, Techniques, and Drug Discovery Applications, in *Tutorial Notes, ACM SIGKDD Int. Conf. On Knowledge Discovery and Data Mining (KDD 2002)*. 2002. Edmonton, Canada.
13. Bohlen, M.H., et al. A Triangular Reconstruction of Density Surfaces, in *Proceedings 3rd International Workshop on Visual Data Mining*. 2003. 19th November, 2003, Melbourne, Florida, USA: University of Technology Sydney.
14. Simoff, S.J., M. Noirhomme-Fraiture, and M.H. Böhlen, eds. *Proceedings of the International Workshop on Visual Data Mining VDM@PKDD2001*. 2001: Freiburg, Germany.
15. Simoff, S.J., M. Noirhomme-Fraiture, and M.H. Böhlen, eds. *Proceedings InternationalWorkshop on Visual Data Mining VDM@ECML/PKDD2002*. 2002: Helsinki, Finland.
16. Simoff, S.J., et al., eds. *Proceedings 3rd International Workshop on Visual Data Mining VDM@ICDM2003*. 2003: Melbourne, Florida, USA.
17. Keim, D.A. and S. Eick, eds. *Proceedings of KDD-2001 Workshop on Visual Data Mining*. 2001: San Francisco, California, USA.
18. Barabasi, A.-L., *Linked: The New Science of Networks*. 2002: Perseus Publishing.

19. Sudweeks, F. and S.J. Simoff, Complementary explorative data analysis: The reconciliation of quantitative and qualitative principles, in *Doing Internet Research*, S. Jones, Editor. 1999, Sage Publications. p. 29-55.
20. Riva, G. and C. Galimberti, Complementary Explorative Multilevel Data Analysis – CEMDA: A socio-cognitive model of data analysis for Internet research, in *Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age*, G. Riva and C. Galimberti, Editors. 2001, IOS Press: Amsterdam. p. 19-35.